

GUIDE FOR THE INTERPRETATION OF THE SIMILARITY REPORT OF THE ANTIPLAGIAT SYSTEM

Similarity Report:

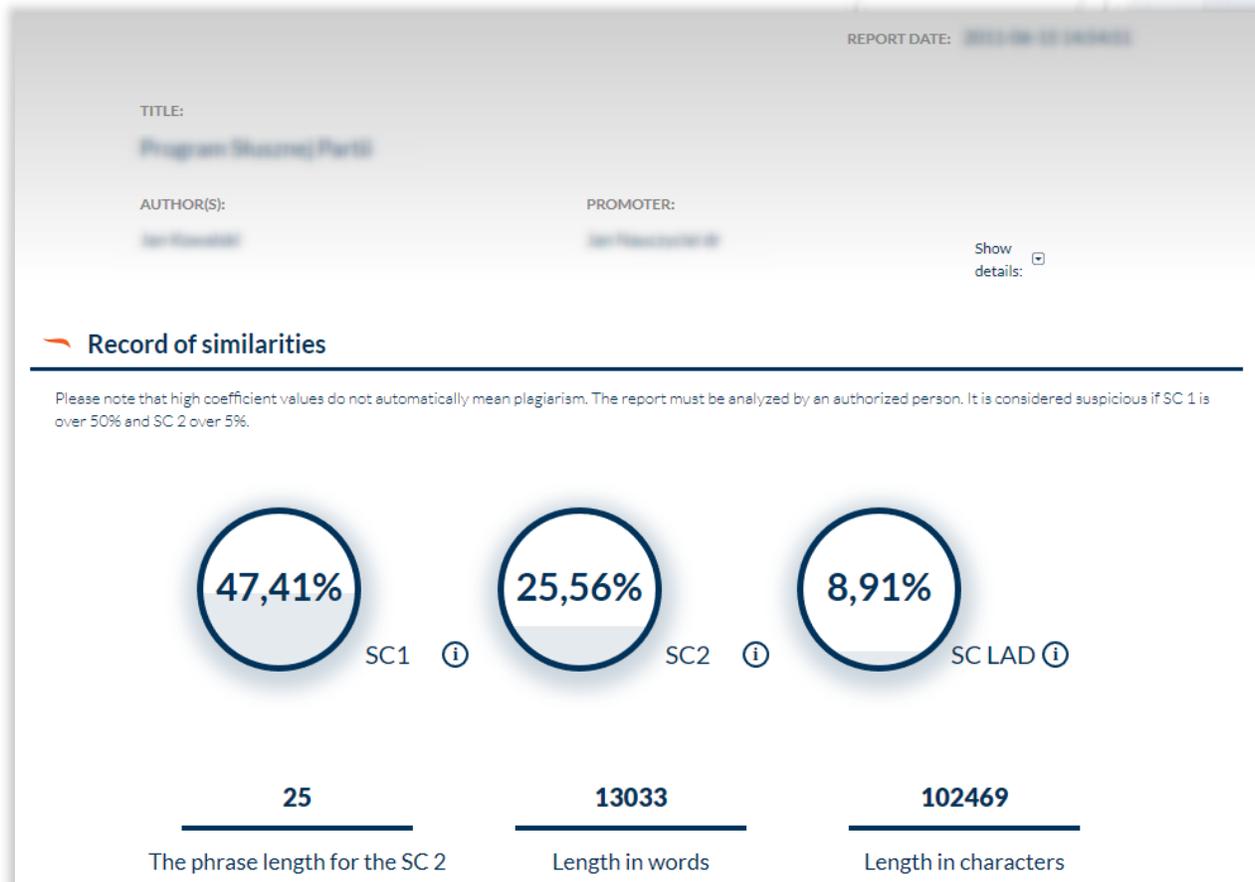
- facilitates the assessment of the independence of the examined text,
- indicates the number of borrowed fragments and gives their source.

I. Functionalities of the Similarity Report

- generates Similarity Coefficients, which determine the percentage of borrowings found in the analyzed document;
- marks the fragments identical to the texts found in the comparative databases,
- presents 10 longest detected in the document borrowing in the form of a list of the longest fragments,
- segregates the borrowings detected in the document according to similarity sources and comparative databases in which they were found (local university database, database of texts collected in the Database Exchange Program, the RefBooks database, the Database of Legal Acts (DLA) and global Internet resources),
- allows distinguishing (by highlighting in blue) selected source of borrowings,
- determines the degree of similarity to a specific similarity source expressed in:
 - number of words common to both texts,
 - number fragments in which it was presented in the analyzed text,
 - the percentage factor of similarity of the analyzed document to the source,
- allows you to navigate through a document facilitating its analysis .

II. Similarity Coefficients

- determine in which percentage the document consists of fragments identical to those detected in other texts,
- express the ratio of the number of words found in other texts to the total number of words in the examined document.



The value of **Similarity Coefficient 1 (SC1)** determines which part of the documents contains phrase of 5 words or longer, found in the database of the university, database of Database Exchange Program, RefBooks database or Internet resources (excluding borrowings from legal acts found in the Database of Legal Acts). The Similarity Coefficient 1 is primarily used to examine the linguistic independence of the author of the documents.

The value of **Similarity Coefficient 2 (SC2)** determines which part of the documents contains phrase of 25 words or longer found in the databases mentioned above (excluding Database of Legal Acts). Due to the length of the detected phrases, the Similarity Score 2 is a better tool for detecting unauthorized borrowings.

Similarity Coefficient DLA (SC DLA) indicates the percentage of the analyzed document that is formed exclusively from phrases of 8 words or longer that are found in the Database of Legal Acts.

III. Active list of similarities

Active list of similarities allows for a quick analysis of the main sources of borrowings in the text and easy navigation on found marked fragments.

Active lists of similarities

Scroll the list and analyze especially the fragments that exceed the SC 2 (marked in bold). Use the link "Mark fragment" and see if they are short phrases scattered in the document (coincidental similarities), numerous short phrases near each other (mosaic plagiarism) or extensive fragments without indicating the source (direct plagiarism).

- + The 10 longest fragments (15,39 %)
- + from RefBooks database (0,00 %)
- + from the Legal Acts Database (8,91 %)
- from the home database (22,53 %)
 All fragments found in the database of your institution.

NO	TITLE	AUTHOR(S)	IDENTICAL WORDS (FRAGMENTS)	CLEAR MARKINGS
1	Fragment 1	Author 1	2 047 (47) 15,71 %	show in the text
2	Fragment 2	Author 2	549 (14) 4,21 %	show in the text
3	Fragment 3	Author 3	340 (16) 2,61 %	show in the text
- + from the Database exchange program (0,00 %)
- + from the Internet (24,88 %)

The percentage value for a specific record in the list indicates what part of the analyzed documents is identical to the indicated source.

IV. Similarity Report Content

The fragments identified as similar have been marked in the Report in colors according to the legend of the origin labels of the borrowings. The colors appearing in the content of the Similarity report mean:

- **green** - fragments from the Internet resources,
- **red** - fragments from the home database and from the Database Exchange Program,

Database Exchange Program - the Antiplagiat system option that allows access to the document databases of other institutions participating in the program, which allows the sources of anti-plagiarism analysis to be expanded.

Membership in the Database Exchange Program results from the provisions of the license agreement. If you are interested in joining the program, please contact the Customer Service Department in order to prepare an annex to the contract and obtain additional information.

- **orange** - fragments from the resources of the RefBooks database,

RefBooks - a database of publications and texts from all fields of science and culture created by Plagiat.pl, constituting the basis for checks performed by the Antiplagiat system. It consists of many collections of texts made available by authors and publishers solely for the purpose of anti-plagiarism analysis.

Currently, it contains nearly 4 million tracks protected by copyright, mainly in Polish and English. Thanks to cooperation with the publishing houses of Wolters Kluwer SA, Termedia and Paperity.org, the database includes the latest books, comments and scientific articles made available both in the traditional model and on the basis of open licenses (the so-called Open Access) and the arxiv.org database. It also includes collections of texts by many smaller publishers and authors.

- **blue background color** - fragments found in documents from the Database of Legal Acts (DLA),

Analyzed texts are also compared with the Polish Legal Acts by Wolters Kluwer SA, which allows to distinguish them from other types of borrowings, because as normative acts are not protected by copyright.

- **yellow background color** - characters from non-Latin alphabet that may indicate attempts to hide borrowings.

Alert - information included in the Similarity Report indicating the occurrence of characters from non-Latin alphabets in the examined document. Alert's function is to pay special attention to the promoter on the legitimacy of using the indicated characters, whose presence may indicate an attempt to falsify the values of the coefficients in the Similarity Report.

pragnąc na zawsze zagwarantować prawa obywatelskie, a działaniu instytucji publicznych zapewnić rzetelność i sprawność,

The documents in which the Alert was identified are highlighted with a red exclamation mark in the document list, and statistics on the alert occurrences are included in the Similarity Report in the "Warnings and Alerts" section.

Warnings and alerts	
Alert	0

- **blue** - fragments of the similarity source selected by the user (using functions located in the drop-down lists).

swiadczen rodzinnych tam, gdzie jest to niezbędne do wyrównania szans dzieci i młodzieży (niepełnosprawność, ubóstwo, dysfunkcjonalność). Musi on promować działania na rzecz usamodzielnienia się rodzin, czyli dostępność pracy i zatrudnienia; upowszechnić usługi opiekuńcze dla małych dzieci (żłobki) i w okresie przed rozpoczęciem edukacji (przedszkola). Istotną rolę w stworzeniu ich sieci muszą odegrać samorządy. One również powinny dbać o jakość tych usług i dostępność finansową; zapewnić młodemu pokoleniu realny dostęp do mieszkań; umożliwić wszystkim dzieciom pełen dostęp do oświaty i nauki oraz różnych dróg edukacyjnych poprzez demokratyczny system stypendialny; Radykalnym zmniejszeniu ubóstwa i wykluczenia wśród dzieci. Można to osiągnąć poprzez: zmiany w zasadach dostępności pomocy (w tym także dożywiania), by docierała ona do dzieci bez względu na postawy rodziców (ich gotowości do współpracy z pracownikami socjalnymi); opracowanie programów wsparcia (o charakterze zadaniowym) skierowanych do rodzin wielodzietnych; upowszechnienie rodzinnych form opieki zastępczej nad dziećmi, zwłaszcza zawodowych rodzin zastępczych; stworzenie warunków do konkurencji na rynku usług opiekuńczych finansowanych przez państwo; wdrożenie efektywnego systemu alimentacyjnego, zapewniającego dzieciom należne im świadczenia.

§ BAP 1. W rozporządzeniu Ministra Finansów z dnia 16 października 2007 r. w sprawie przyznania Agencji Restrukturyzacji i Modernizacji Rolnictwa tymczasowej akredytacji jako agencji płatniczej, w zakresie uruchamiania środków pochodzących

LEGEND

- - Institutional Databases
- - Internet Sources
- - legal acts
- - RefBooks Database

Two shades of green and red are used to distinguish the following fragments found in different sources or in the same source, but in other places.

V. Additional functions

On request may be included in the additional functionality of the Report:

a. Quotes Coefficient

Value of **Quotes Coefficient (QT)** indicates what part of the analyzed document has been included in quotes (i.e. Begins and ends with „ ”).



It should be remembered that the system will only select quotes correctly marked with the quotation marks. The system does not analyze the legitimacy of using fragments covered by the quotations. For this reason, quotes are not omitted during the antiplagiarism analysis. The Quotes Coefficient is ancillary and the mechanical nature of its calculation should be taken into account when analyzing the results of the report.

. Słowo „cytat” jako określenie dopuszczalnego przytoczenia wcześniejszego tekstu upowszechniło się w językach europejskich na początku Oświecenia. W 1812, w okresie powstawania pierwszych ustaw autorskoprawnych, Charles Nodier napisał, że „nie można usprawiedliwiać żadnych zapożyczeń z wcześniejszych dzieł, z wyjątkiem cytatów” [6]

b. Skipping the bibliography

It is possible to start omitting the analysis of the bibliography contained in the work, provided that it was presented in the form of a numbered list beginning with the word "Bibliography".

Polska, Rzeczpospolita Polska (RP) – państwo unitarne w Europie Środkowej, położone między Morzem Bałtyckim na północy a Sudetami i Karpatami na południu, w przeważającej części w dorzeczu Wisły i Odry. Od północy Polska graniczy z Rosją (z jej obwodem kaliningradzkim) i Litwą, od wschodu z Białorusią i Ukrainą, od południa ze Słowacją i Czechami, od zachodu z Niemcami. Większość północnej granicy Polski wyznacza wybrzeże Morza Bałtyckiego. Polska Wyłączna Strefa Ekonomiczna na Bałtyku graniczy ze strefami Danii i Szwecji.

Bibliografia

1. Powierzchnia i ludność w przekroju terytorialnym w 2011 r., s. 17, 2011. Departament Metodologii, Standardów i Rejestrów. GUS. [dostęp 2015-07-19].
2. Halina (red.) Dmochowska. Warunki naturalne i ochrona środowiska. „Mały Rocznik Statystyczny Polski 2015”, s. 17–18, 2011-07-21. Warszawa: Główny Urząd Statystyczny. ISSN 1640-3630. [dostęp 2015-07-19].
3. Główny Urząd Statystyczny. Ludność. Stan i struktura oraz ruch naturalny w przekroju terytorialnym w 2017 r. Stan w dniu 31 XII [1].

Powierzchnia administracyjna Polski wynosi 312 679 km²[1], co daje jej 69. miejsce na świecie[6] i 9. w Europie. Zamieszkała przez prawie 38,5 miliona ludzi (2017), zajmuje pod względem liczby ludności 36. miejsce na świecie[7], a 6. w Unii Europejskiej. Polska podzielona jest na 16 województw. Jej największym miastem i jednocześnie stolicą jest Warszawa. Inne metropolie to Kraków, Łódź, Wrocław, Poznań, Gdańsk, Szczecin. Największą polską aglomeracją policentryczną jest konurbacja górnośląska.

VI. Interpretation of the Similarity Report

To make a precise interpretation of the Similarity Report, the following steps must be taken:

- Determine the values for the similarity coefficients (it is considered suspicious if the first Similarity Coefficient exceeds 50% and if the second Similarity Coefficient exceeds 5%);
- Examine the list "Longest Fragments Identified as Similar" (fragments which have a number of words bigger than 50 are considered suspicious, therefore they require a rigorous verification; if this kind of fragment exists, it must be found by using the link "Mark fragment" and to verify whether it is cited or not);
- Scroll the lists "Documents Containing Similar Fragments", taking into account especially the documents containing fragments that exceed the Similarity Coefficient 2 (these are marked in bold). In the you have such documents, and especially if they are at the top of the list, you must use the link "Mark fragment" and verify whether they are short phrases scattered throughout the document (in this case we can consider them as coincidental loans) or long text fragments that are adjacent and separated only by short phrases (this kind of situation will arouse suspicion);
- If any suspicions arise, the Similarity Report should be subjected to an accurate analysis, which is based, in addition to using the Similarity Report functions presented in paragraphs 1-3, on the exact analysis of the content of the document, taking into consideration the fragments that were discovered by the system in other texts.

VII. Basic information about the interpretation of the Similarity Report

Plagiat.pl is a tool for verifying the originality of checked documents. Its purpose is to determine the exact proportion of the possible similarities of the verified text in comparison to the content of the databases and the Internet.

The system provides the above information, allowing an independent assessment on the legitimacy of borrowings found in the verified document. The purpose of the system is not to declare if the text was written independently or not, but to provide the needed materials in order to form an opinion regarding its originality. Therefore, the Similarity Report should always be examined by a competent person. In particular, a document must not be assessed based solely of the percentages of the Similarity Coefficients. It is necessary to check the documents content - if quotations are marked and if they come from documents listed under references.

Plagiat.pl does not determine which document was created first - the one analyzed by the system or the one found that is considered source. In case of doubts, the user can not establish, solely based on the Similarity Report, which of the documents is the original and which was copied. This conclusion can only result from a detailed analysis of both documents.

Due to the methods used to analyze similarities, the system will also detect certain phrases that are often used, such as: "As I have mentioned before" or "We are able to conclude that". However, the number of such phrases in the group of loans and their impact on the percentage similitude coefficient value should not exceed 50%.

The analysis conducted by our company show that when texts contain a large number of phrases taken from a professional field, similarity coefficient increases significantly. As a result, some documents may obtain relatively high percentages, although a detailed analysis of the Similarity Report will show that it does not contain unauthorized loans. For this reason we introduced the second similarity coefficient. Similarity Coefficient 2 defines a more accurate percentage of loans found in the analyzed documents. Its value is calculated as the Similarity Coefficient 1, but it identifies phrases that have a certain number of words, which is set by the university. Our recommendation for the universities we work with is to set the limit for Similarity Coefficient 2 up to 25 words. Also, we recommend a detailed analysis of reports which have Similarity Coefficient 2 more than 5%. Documents that exceed permissible coefficients of similarity percentages cannot be automatically considered cases of plagiarism.

VIII. Contact:

If you have any problems, please contact our Customer Service Department,

E-mail: pomoc@antyplagiat.pl

Phone: (22) 100 11 11